Journal of Experimental Botany, Vol. 64, No. 7, pp. 1983–1994, 2013
doi:10.1093/jxb/ert057 Advance Access publication 25 March, 2013
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Generalized additive models reveal the intrinsic complexity of wood formation dynamics

Henri E. Cuny[1,*], Cyrille B.K. Rathgeber[1], Tristan Senga Kiessé[1], Felix P. Hartmann[1], Ignacio Barbeito[1] and Meriem Fournier[2]

[1] INRA, UMR1092, Laboratoire d Etude des Ressources Foret Bois (LERFoB), Centre INRA de Nancy, F-54280 Champenoux, France
[2] AgroParisTech, UMR1092, Laboratoire d Etude des Ressources Foret Bois (LERFoB), ENGREF, 14 rue Girardet, F-54000 Nancy, France

*To whom correspondence should be addressed. E-mail: henri.cuny@nancy.inra.fr

Received 28 December 2012; Revised 8 February 2013; Accepted 14 February 2013

Abstract

The intra-annual dynamics of wood formation, which involves the passage of newly produced cells through three successive differentiation phases (division, enlargement, and wall thickening) to reach the final functional mature state, has traditionally been described in conifers as three delayed bell-shaped curves followed by an S-shaped curve. Here the classical view represented by the 'Gompertz function (GF) approach' was challenged using two novel approaches based on parametric generalized linear models (GLMs) and 'data-driven' generalized additive models (GAMs). These three approaches (GFs, GLMs, and GAMs) were used to describe seasonal changes in cell numbers in each of the xylem differentiation phases and to calculate the timing of cell development in three conifer species [*Picea abies* (L.), *Pinus sylvestris* L., and *Abies alba* Mill.]. GAMs outperformed GFs and GLMs in describing intra-annual wood formation dynamics, showing two left-skewed bell-shaped curves for division and enlargement, and a right-skewed bimodal curve for thickening. Cell residence times progressively decreased through the season for enlargement, whilst increasing late but rapidly for thickening. These patterns match changes in cell anatomical features within a tree ring, which allows the separation of earlywood and latewood into two distinct cell populations. A novel statistical approach is presented which renews our understanding of xylogenesis, a dynamic biological process in which the rate of cell production interplays with cell residence times in each developmental phase to create complex seasonal patterns.

Key words: Cambial activity, conifers, generalized linear and additive models (GLMs and GAMs), Gompertz functions (GFs), timing of cell development, tree ring, wood formation, xylogenesis.

Introduction

Despite the key role of wood in terrestrial biosphere and human societies, our understanding of the complex biological processes involved in its formation remains fragmentary (Plomion *et al.*, 2001; Chaffey, 2002). In particular, the dynamics (timing, duration, and rate) of the wood formation phases (cambial division, cell enlargement, and secondary cell wall formation) remains poorly explored, despite this being of fundamental importance in determining the quantity (number of cells produced) and quality (cell anatomical characteristics) of the wood formed (Vaganov *et al.*, 2011).

In recent years, a growing number of studies have been devoted to wood formation monitoring, and these have provided excellent data sets (see, for example, Rossi *et al.*, 2006*b*, 2008). However, most of these studies have only focused on key aspects of the process, such as the onset and cessation of xylem formation (Rossi *et al.*, 2011), the maximal rate of cell production (Rossi *et al.*, 2006*b*), or more general intra-annual

growth patterns (Camarero *et al.*, 2010; Rathgeber *et al.*, 2011; Cuny *et al.*, 2012). Thus, a complete description of wood formation dynamics can only be found in isolated studies (Wodzicki, 1971; Horacek *et al.*, 1999).

During wood formation, cells successively pass through three phases to reach a permanently mature state: (i) division; (ii) enlargement; and (iii) wall thickening (Wilson *et al.*, 1966). The temporal succession of the differentiation phases undergone by the cells forms a spatial succession of developmental zones along the tree ring. Drawing inspiration from hydrological system modelling, the three developmental zones and the mature zone can be viewed as four interconnected reservoirs, with cells 'flowing' from one reservoir to the next as they differentiate (Supplementary Fig. S1A available at *JXB* online). In conifers, the seasonal dynamics of the number of cells passing through the successive developmental zones has traditionally been described as three delayed bell-shaped curves followed by an S-shaped curve (Supplementary Fig. S1B) (Rossi *et al.*, 2006*a*; Lupi *et al.*, 2010; Cuny *et al.*, 2012). This classical view is based on the assumption that the flow of the source (i.e. the rate of cell production) follows a bell-shaped curve through the season (Larson, 1994) and drives the dynamics of the subsequent reservoirs.

Through the growing season, cells accumulate asymmetrically, with an initial positive exponential phase followed by gradual slowing. Consequently, the Gompertz function (GF)—an asymmetric sigmoid growth function, where the rate is maximal at one-third of the asymptote before to slow down progressively—has been used to fit the data and calculate the timing of cell development (Supplementary Fig. S1C at *JXB* online) (Camarero *et al.*, 1998; Rossi *et al.*, 2003). This method allows variations in cell residence times in the different zones of xylogenesis to be calculated through the season, but only where these are close to linear (Rossi *et al.*, 2006*a*; Deslauriers *et al.*, 2003).

Some pioneer studies have found that developing tracheids spend less time in enlargement and more time in thickening as the growing season proceeds (Whitmore and Zahner, 1966; Skene, 1969, 1972; Wodzicki, 1971; Kutscha *et al.*, 1975). Such findings argue against the idea of repeated bell-shaped curves, which implies that cell residence times change in parallel between the successive reservoirs through the season. Instead, they suggest that developing tracheids accumulate in the enlargement reservoir at the beginning of the season, when the increase in cell diameter is greatest, and then accumulate in the thickening reservoir at the end of the season, when the main focus is on building and lignifying their thick cell walls.

This led to the proposal of the following hypothesis: the rate of cell transition from one development zone to the next and the cell residence times in these zones are not monotonous through a growing season, so that the resulting changes in the number of cells in the successive zones should be intrinsically more complex than the classical view; additionally, it should be possible to see the footprint of cell anatomical features in the intra-annual dynamics of the process that produced them, so that we can distinguish the transition between earlywood (large, thin-walled cells) and latewood (small, thick-walled cells) in the intra-annual pattern of enlarging and thickening cell numbers.

To test this hypothesis, three statistical modelling approaches were applied to three conifer species (to ensure robustness) over 3 years of monitoring (to ensure generality): (i) the traditionally used parametric GF, which fits bell- and S-shaped distributions and represents the classical view; (ii) another parametric approach using generalized linear models (GLMs); and (iii) a semi-parametric approach using generalized additive models (GAMs), which are more flexible and are thus able to highlight structures in the data distribution that differ from bell- or S-shaped curves, thereby allowing the classical view of wood formation dynamics to be challenged.

Materials and methods

Study site, xylem sampling, and sample preparation

Dominant and healthy Scots pines (*Pinus sylvestris* L.), silver firs (*Abies alba* Mill.), and Norway spruces [*Picea abies* (L.) Karst.] (n=5 per species) were selected in a mixed stand located in the Vosges Mountains (48°29'N, 7°09'E, and 643 m a.s.l.), in northeast France (Supplementary Table S1 at *JXB* online). Microcores were collected weekly from the tree stems from April to November 2007–2009 and prepared in the laboratory to obtain transverse sections (5–10 µm thick), which were stained with cresyl violet acetate and permanently mounted on glass slides. For detailed information about the study site, tree characteristics, sampling strategy, and sample preparation, see Cuny *et al.* (2012).

Microscopic observations of xylem cell differentiation

Overall, 1450 anatomical sections were observed using an optical microscope (AxioImager.M2; Carl Zeiss SAS, France) under visible and polarized light at ×100–400 magnification to distinguish the cells in the different developmental zones. Cambial cells had a rectangular shape, were surrounded by a thin primary wall, and had small radial diameters, whereas cells in the enlargement zone still had thin primary walls but had larger radial diameters. In contrast to cells in the cambial and enlargement zones, cells in the thickening zone developed a secondary wall that appeared bright under polarized light because of its particular arrangement of cellulose microfibrils (Abe *et al.*, 1997). Cresyl violet acetate staining, whereby cellulose stains purple and lignin stains blue (Kutscha *et al.*, 1975), was used to follow the advancement of lignification, with cells in the thickening phase exhibiting purple and blue walls, indicating that lignification was in progress, whereas mature cells had entirely lignified and thus had completely blue walls.

Gompertz function, and generalized linear and additive models

For each sample, the radial number of cells in the cambial (n_C, see Supplementary Table S2 at *JXB* online for the list of variables used along with their notations and units), enlargement (n_E), thickening (n_T), and mature (n_M) zones was counted along three radial cell files (see Supplementary Table S3 for a description of the number of cells in each zone and their variation). The number of cells from the previous year was also counted on three radial files per sample and used to standardize the number of cells of the current year to reduce within-tree growth variability (Rossi *et al.*, 2003). A dedicated function of the package CAVIAR (Rathgeber, 2012) was used to apply this standardization using R statistical software (R Development Core Team, 2011).

Following Rossi *et al.* (2003), GFs were fitted to the standardized number of mature cells, which is cumulative.

Thus, the GF is defined as:

$$N(t) = A.e^{-e^{\beta-\kappa.t}} \quad (1)$$

where $N(t)$ is the number of mature cells at time t; A is the upper asymptote parameter representing the final number of cells; β is the x-axis placement parameter that reflects the location of the origin; and κ is the growth rate parameter that determines the spread of the curve along the time axis.

However, in contrast to previous studies (Rossi *et al.*, 2003), here it was considered a better practice to fit the first derivative of the Gompertz function (GF') directly to the standardized number of cells in the division, enlargement, and thickening zones.

Thus, GF' is expressed as:

$$N(t) = A.\kappa.e^{-e^{\beta-\kappa.t}}.e^{\beta-\kappa.t} \quad (2)$$

GF and GF' parameters were estimated using the nls function in R, which determines the non-linear (weighted) least-squares estimates of the parameters of a non-linear model (Bates and Chambers, 1992).

Additionally, two new approaches were used to model intra-annual wood formation dynamics based on GLMs and GAMs. The advance in regression analysis provided by GLMs and GAMs has been an important statistical development over the last 30 years. GAMs are semi-parametric extensions of GLMs, which are themselves mathematical extensions of classical linear models (Wood, 2006).

The advantages of GLMs are that they define (i) a link function between a non-linear response variable and explanatory variables; and (ii) a probability distribution for non-normal errors (McCullagh and Nelder, 1983). This makes them particularly appropriate for modelling count data (in our case cell number variations) by using a log as a link function, and a Poisson distribution for errors.

GAMs are GLMs in which the linear predictor depends, in part, on a sum of smooth functions of the predictors. GAMs are referred to as being data driven rather than model driven, because the data determine the nature of the relationship between the response variable and the set of explanatory variables rather than assuming a certain type of parametric relationship (Hastie and Tibshirani, 1986). The strength of GAMs lies in their ability to deal with highly non-linear and non-monotonic relationships between the response and the set of explanatory variables, which helps develop ecological models that better represent the underlying data, thereby increasing our understanding of biological systems (Guisan *et al.*, 2002).

In this study, the weekly count of cells in each zone of wood formation was expressed as a function of the day of the year:

$$\log(n)=\alpha+s(d)+\varepsilon \quad (3)$$

where n is the vector of the weekly count of cells in the considered zone, d is the vector of the corresponding day of the year, s is an unspecified smooth function, α is the intercept, and ε is the error term. GAMs were fitted in R using the mgcv package (Wood, 2006).

Gompertz functions (GFs and GFs′), GLMs, and GAMs were fitted for every year on each individual tree. The values of the fitted models were then averaged over the three studied years in order to calculate means representing the general wood formation dynamics of a species.

Goodness-of-fit measurement

The goodness-of-fit of the GFs, GLMs, and GAMs was assessed by computing the mean absolute error (MAE), as this is the most natural and most easily interpretable measure of model accuracy (Willmott and Matsuura, 2005). The MAE gives the mean absolute difference between the model predicted values and the observations in the units of the response:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|}{n} \quad (4)$$

where y_i is the ith observed value, and $\hat{y}_i$ is the corresponding model-fitted value.

One problem with the MAE is that it depends on the scale of the response, meaning that it cannot be compared across series. Therefore, to compare the fittings between the different zones, in which the cell numbers differed, the mean absolute percentage error (MAPE) was also computed (Armstrong and Collopy, 1992):

$$\mathrm{MAPE} = \sum_{i=1}^{n}\frac{\left|\frac{y_i - \hat{y}_i}{\bar{y}}\right|}{n} \quad (5)$$

where y_i is the ith observed value, $\hat{y}_i$ is the corresponding fitted value, and $\bar{y}$ is the mean of the n observed values.

In addition, the modelling efficiency (EF), a relative measure of the goodness-of-fit that gives the proportion of variation in the response captured by the model, was calculated (Mayer and Butler, 1993). This statistic is similar to the coefficient of determination (r^2), but can be used for non-linear models:

$$EF = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (6)$$

EF is close to 1 when there is a good fit, and is ≤0 when the model has equal or lower predictive power than the mean of the observations.

The Akaike information criterion (AIC), which penalized the goodness-of-fit of the model by the number of parameters, was also calculated (Akaike, 1973):

$$\mathrm{AIC} = -2.\log(L) + 2.k \quad (7)$$

where L refers to the maximized value of the likelihood function for the fitted model and k is the number of parameters. AIC is a relative measure: it tells nothing about how well a model fits the data in an absolute sense, but is used to compare models. In a set of models, the model that has the lowest AIC is selected because it provides the best compromise between the goodness-of-fit and the number of parameters.

Residence duration

Four numerical functions (S) were defined to represent the cumulative number of cells:

$$S_{CETM}(t) = n_C(t) + n_E(t) + n_T(t) + n_M(t) \quad (8)$$

$$S_{ETM}(t) = n_E(t) + n_T(t) + n_M(t) \quad (9)$$

$$S_{TM}(t) = n_T(t) + n_M(t) \quad (10)$$

$$S_M(t) = n_M(t) \quad (11)$$

where n represents functions (GFs, GLMs, or GAMs) that predict the number of cells in the cambial (n_C), enlargement (n_E), thickening

(n_T), and mature (n_M) zones at date t. Because these numerical functions are strictly increasing, their inverse function (S^{-1}) can be used to compute cell timings. For a cell i, the dates of entrance into the enlargement ($t_{E,i}$), thickening ($t_{T,i}$), and mature ($t_{M,i}$) zones were computed as:

$$t_{E,i} = S_{ETM}^{-1}(i) \quad (12)$$

$$t_{T,i} = S_{TM}^{-1}(i) \quad (13)$$

$$t_{M,i} = S_{M}^{-1}(i) \quad (14)$$

For each approach, the time spent by each cell i in the enlargement ($d_{E,i}$) and thickening ($d_{T,i}$) zones, and the total duration of cell formation ($d_{F,i}$) were calculated as:

$$d_{E,i} = t_{Ti} - t_{Ei} \quad (15)$$

$$d_{T,i} = t_{Mi} - t_{Ti} \quad (16)$$

$$d_{F,i} = t_{Mi} - t_{Ei} \quad (17)$$

Transition rate

For each approach, the rate of cell production by the cambial zone ($r_{C,t}$) and the entry rate of cells into the enlargement ($e_{E,t}$), thickening ($e_{T,t}$), and mature ($e_{M,t}$) zones at day t were computed as:

$$r_{C,t} = S_{CETM}(t) - S_{CETM}(t-1) \quad (18)$$

$$e_{E,t} = S_{ETM}(t) - S_{ETM}(t-1) \quad (19)$$

$$e_{T,t} = S_{TM}(t) - S_{TM}(t-1) \quad (20)$$

$$e_{M,t} = S_{M}(t) - S_{M}(t-1) \quad (21)$$

Results

Comparative performance of models

GAMs gave the best fits to the changes in cell numbers in the different zones of xylogenesis, with errors (MAE and MAPE) being, on average, twice as low as those computed for GLM and GF fittings (Table 1). Model efficiencies were also, on average, 30% higher for GAMs than for GLMs and GFs, showing that GAMs explained cell number variations better than the other models. Finally, AICs computed from GAMs were 30% lower than those computed from GF and GLM, indicating that GAMs offer the best trade-off between goodness-of-fit and parsimony.

Changes in cell numbers in the zones of xylogenesis

The number of cambial cells increased rapidly and culminated early, at the beginning of June, for all three species (Fig. 1A–C). This remarkable pattern was well reproduced by the GAMs, which showed bell-shaped curves skewed to the left (Fig. 2A–C). In contrast, GLMs and GFs were not able to preserve this pattern, delaying the occurrence of maximal cell numbers until later in the season and underestimating them (Table 2). Later in the season, the number of cambial cells gradually decreased and finally returned to approximately its initial value. Once again, GAMs reproduced this progressive decrease better than GFs and GLMs.

In the enlargement zone, the number of cells also increased rapidly at the beginning of the growing season, in April. The number culminated at the very beginning of June, before decreasing progressively back to zero in September–October, indicating the end of tree growth (Fig. 1D–F). All three approaches were able to capture this pattern, showing bell-shaped curves skewed to the left (Fig. 2D–F), with curves obtained using GFs being close to those obtained using GAMs. GLMs, however, were unable to follow the rapid early increase in enlarging cells, again resulting in a shift of the maximum to later in the season (Table 2).

In the thickening zone, the number of cells rapidly increased and reached an initial peak in the first half of June (Fig. 1G–I). The number then plateaued or slightly declined over 2–3 weeks, and then increased again to reach a second peak later in the season, which was greater than the first and thus corresponded to the maximum number of cells observed in this zone. In spruce, this maximum occurred at the beginning of August, which was 2 weeks earlier than in fir (mid-August) and 1 month earlier than in pine (beginning of September). The difference in magnitude between the two peaks differed between species, with the second late peak being 30% greater than the first peak in fir and spruce, but 100% greater in pine. Following this late maximum, the number of cells rapidly decreased until November, when no more cells remained in the thickening zone, indicating the end of biomass allocation to the stem. Only GAMs were able to capture the complex pattern of cell number changes in the thickening zone, clearly showing a bimodal curve skewed to the right (Fig. 2G–I). In contrast, the less flexible GFs and GLMs presented bell-shaped curves that exhibited their unique maximum at the place of the local minimum between the early and late peaks indicated by the raw data and GAMs (Figs 1G–I, 2G–I).

From the beginning of June to November, there was a steady accumulation of cells in the mature zone (Fig. 1J–L). Only GAMs were able to detect two periods of faster accumulation separated by a period of slower increase (Fig. 2J–L).

Cell development duration

All three modelling approaches allowed calculation of the time a particular tracheid spends in each zone of xylogenesis during the course of its differentiation (Supplementary Fig. S2 at *JXB* online). All methods agreed on the general trends in cell development duration through a tree ring, showing that cells spent less time in the enlargement zone (d_E) and more time in the thickening zone (d_T) from the beginning to the end of a ring (Fig. 3). Furthermore, since d_E was always less than or equal to d_T, the total duration of tracheid formation (d_F) increased from the beginning to the end of a ring.

However, a more detailed look showed that the three approaches presented contrasting results. GFs and GLMs, which returned similar curves when fitting cell numbers in the different zones of xylogenesis, exhibited contrasting

Table 1. Mean absolute error (MAE), mean absolute percentage error (MAPE), model efficiency (EF), and Akaike information criterion (AIC) computed from the fittings of the Gompertz functions (GFs), and generalized linear and additive models (GLMs and GAMs) on the number of cells in the cambial (CZ), diameter enlargement (EZ), wall thickening (TZ), and mature (MZ) zones for Scots pine, Norway spruce, and silver fir (n=5 per species) during 2007, 2008, and 2009 (means).

Zone	Species	MAE (cell)			MAPE (%)			EF (%)			AIC		
		GF	GLM	GAM	GF	GLM	GAM	GF	GLM	GAM	GF	GLM	GAM
CZ	Pine	0.75	0.74	0.39	11	11	6	43	45	83	246	248	227
	Fir	0.83	0.82	0.55	9	10	7	29	30	61	247	249	232
	Spruce	0.73	0.71	0.39	11	11	6	50	52	82	245	245	228
EZ	Pine	0.61	0.68	0.28	32	35	14	81	73	94	226	256	160
	Fir	0.81	0.88	0.31	44	46	17	68	59	93	254	333	164
	Spruce	0.41	0.46	0.17	35	38	15	83	75	95	227	198	135
TZ	Pine	1.72	1.73	0.71	34	34	14	67	67	92	300	519	236
	Fir	3.02	3.08	1.06	30	30	11	74	74	96	338	805	292
	Spruce	1.48	1.4	0.45	36	33	10	78	78	97	293	456	198
MZ	Pine	1.64	1.25	1.01	14	11	9	94	95	97	302	320	238
	Fir	2.85	2.24	1.44	11	9	5	97	97	99	330	442	271
	Spruce	1.47	1.39	0.84	12	11	7	96	96	98	295	312	217

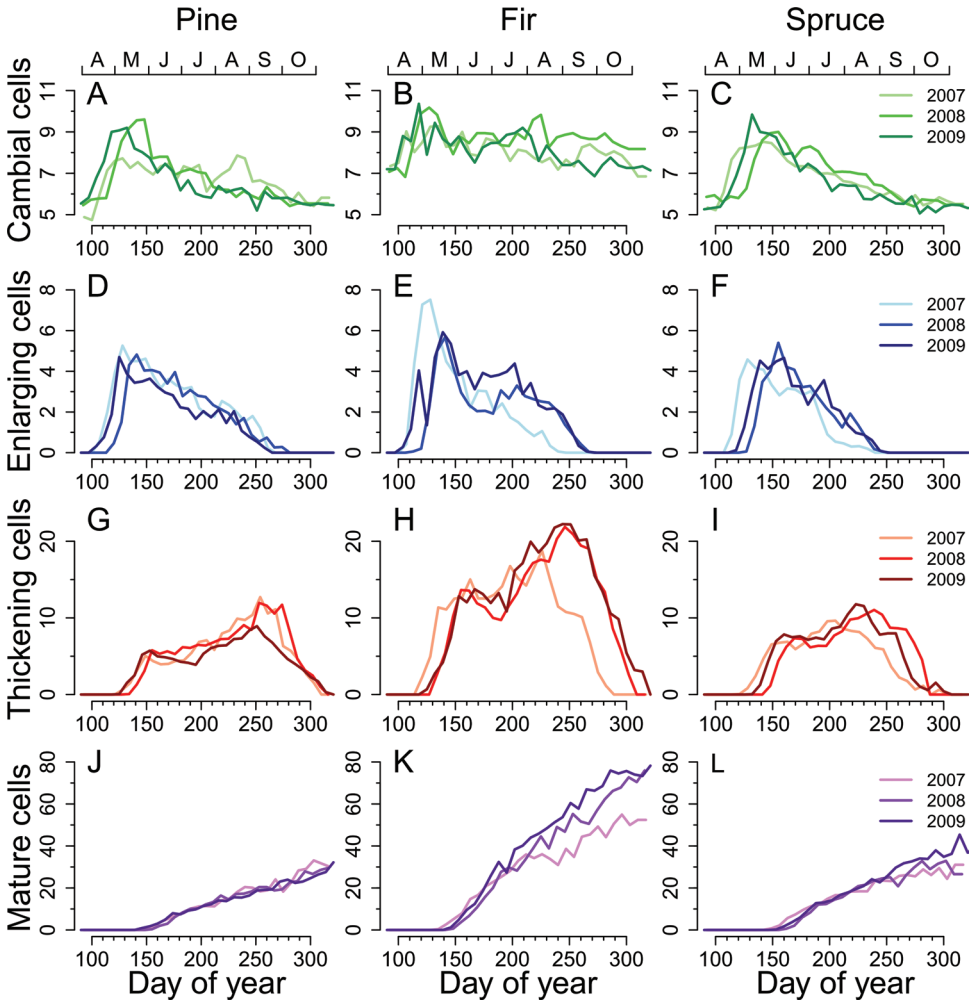

Fig. 1. Number of cells in the cambial, enlargement, thickening, and mature zones of wood formation for Scots pine, silver fir, and Norway spruce. Lines represent the mean of five trees per species per year. The upper *x*-axis represents the months.

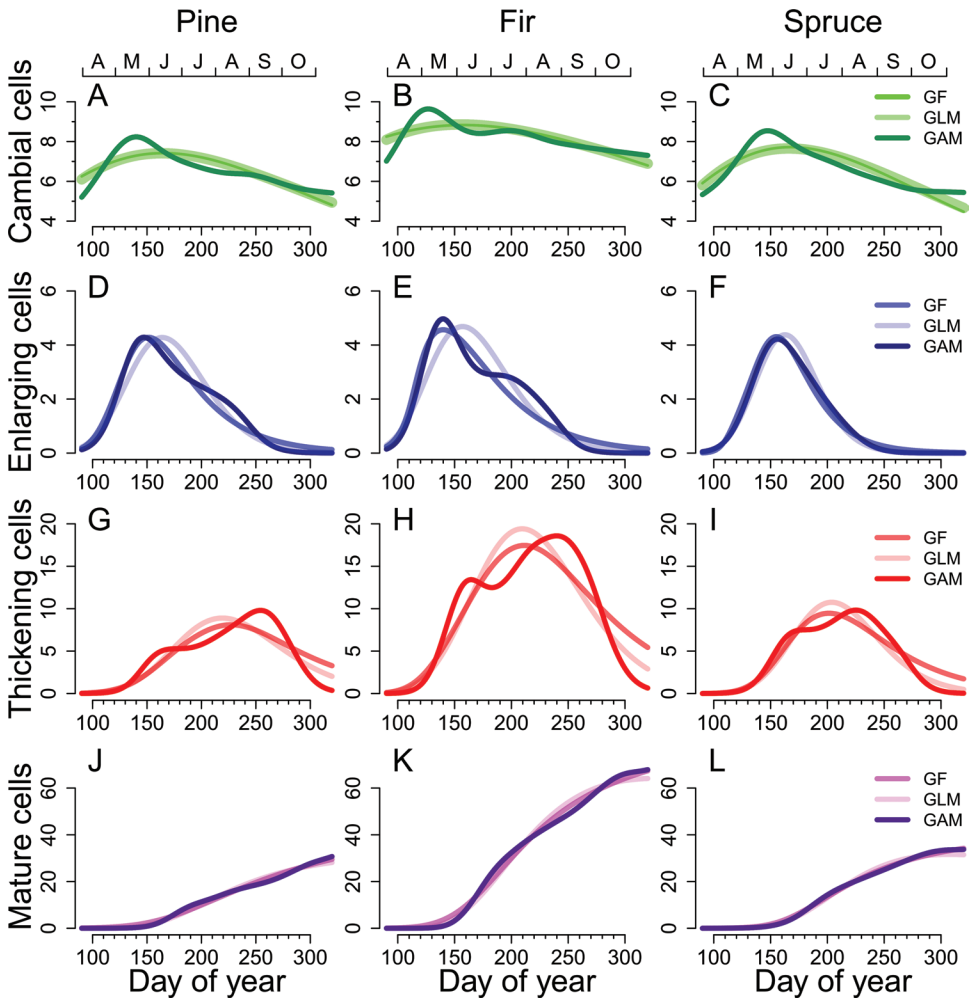

Fig. 2. Gompertz functions (GFs), and generalized linear and additive models (GLMs and GAMs) fitted to the weekly cell count in the cambial, enlargement, thickening, and mature zones for Scots pine, silver fir, and Norway spruce (*n*=5 per species) over 3 years (2007, 2008, and 2009). The upper *x*-axis represents the months.

monotonous patterns, which were always convex for GFs and always concave for GLMs. In contrast, GAMs gave an accurate fit, capturing realistic patterns of cell development durations. The lack of fit of GFs and GLMs resulted in a systematic underestimation of cell development duration (30% and 10%, respectively) compared with GAMs.

Table 2. Maximum number (n_x) and date of occurrence of the maximum number (t_x) of cells in the cambial (CZ), diameter enlargement (EZ), and wall thickening (TZ) zones of wood formation, computed from the fittings of the Gompertz functions (GFs), and generalized linear and additive models (GLMs and GAMs) for Scots pine, Norway spruce, and silver fir (n=5 per species) during 2007, 2008, and 2009 (means ±SEs).

Zone	Species	n_x (cell)			t_x (day of year)		
		GF	GLM	GAM	GF	GLM	GAM
CZ	Pine	7.6±0.3	7.7±0.3	8.5±0.3	158±9	156±8	148±10
	Fir	9.1±0.3	9.1±0.3	10.0±0.5	162±15	161±15	148±14
	Spruce	7.8±0.2	7.7±0.2	8.8±0.3	163±6	162±4	151±3
EZ	Pine	4.5±0.3	4.4±0.3	4.5±0.3	153±2	164±2	151±4
	Fir	5.3±0.5	5.1±0.4	5.5±0.5	150±4	161±3	149±6
	Spruce	4.8±0.3	4.7±0.3	4.6±0.3	152±3	162±3	154±2
TZ	Pine	8.7±0.9	9.2±0.8	10.7±1.2	218±4	218±3	247±6
	Fir	18.8±1.8	20.6±2.1	21.3±2.1	208±5	211±4	229±9
	Spruce	10.2±0.9	11.6±1.0	11.4±1.2	198±4	202±4	216±7

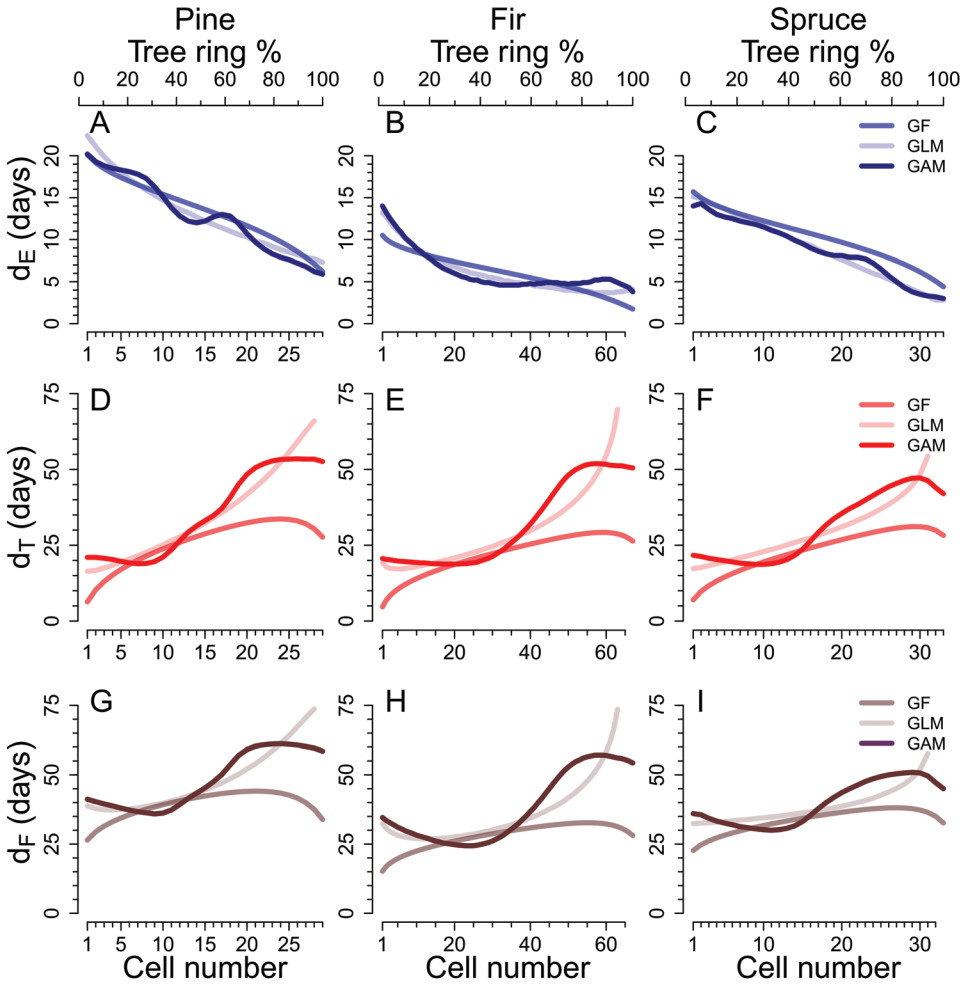

Fig. 3. Cell residence durations in the enlargement (d_E) and thickening (d_T) zones, and total duration of formation (d_F) of tracheids along tree rings in Scots pine, silver fir, and Norway spruce as computed from Gompertz functions (GFs), and generalized linear and additive models (GLMs and GAMs).

The GAMs showed that d_E was longest in pine (13 ± 4 d), ~30% shorter in spruce (9 ± 3 d), and 50% shorter in fir (6 ± 2 d). d_E was also maximal for the first-formed cells of the year, ~20 d in pine, and 14 d in fir and spruce (Fig. 3). It decreased by approximately two-thirds along the tree ring (with final durations of 6 d in pine, 5 d in fir, and 4 d in spruce). However, while d_E decreased progressively in pine and spruce, in fir it decreased rapidly in the first 30% of cells, and then remained stable around its minimal value for the following cells. The GAMs also showed that the ring could be separated into three distinct groups of cells according to d_T (Fig. 3D–F). The first group corresponded to the first 40% of cells, for which d_T remained at a steady minimum of ~20 d (regardless of species). The second group corresponded to the middle 40% of cells, which showed a strong increase in d_T; the first cells in this second group entered thickening at the end of June, when the number of cells in this zone started to increase again (Fig. 2). The third group corresponded to the last 20% of cells, which had the maximal d_T (53 ± 1 d in pine, 51 ± 1 d in fir, and 45 ± 2 d in spruce); the cells belonging to this third group entered thickening in August and/or September, when the number of cells in this zone was maximal. On average, d_T was 10% shorter in fir and spruce than in pine (32 ± 14, 31 ± 11, and 36 ± 14 d, respectively).

The longer d_E and d_T in pine cells meant that d_F was also 20% longer in pine than in fir and spruce (49 ± 10, 38 ± 13, and 40 ± 8 d, respectively). d_F followed a complex pattern resulting from the combination of d_E and d_T. In the first 40% of cells, d_F decreased slightly (from 41 to 36 d in pine, 37 to 30 d in spruce, and 35 to 24 d in fir). It then increased rapidly to reach its maximum at ~80% of a ring (61 d in pine, 57 d in fir, and 51 d in spruce). In the last 20% of cells, d_F decreased slightly for a few days.

On average, 20% of d_F was explained by d_E and 80% by d_T, with little variation between species in these proportions (15% versus 85% in fir, 20% versus 80% in spruce, and 25% versus 75% in pine). However, these relative contributions varied greatly according to the position of the cells along the ring: the relative contribution of d_E to d_F in the first 40% of cells was ~50% in pine, 40% in spruce, and 30% in fir, but this fell to ~10% for the last 20% of cells in a ring, regardless of species.

Entry rates into the different zones

Cell entry rates into the successive zones of xylogenesis exhibited more complex seasonal patterns when calculated using GAMs than the classic bell-shaped curves depicted when using GFs or GLMs (Fig. 4). The cell division rates (and so also the entry rates) depended on the productivity of the species studied, with fir reaching ~0.7 cells d^{-1}, spruce reaching 0.35 cells d^{-1}, and pine remaining at ~0.25 cells d^{-1}.

Despite these contrasting levels, the three species presented similar changes in their rates of entry into the different zones of xylogenesis during the growing period. Moreover, the division rate, and the enlargement and thickening entry rates followed close to parallel patterns, only being staggered by a few days. These three rates presented a bimodal pattern: the first peak occurred at the end of May, around the time when the number of cells culminated in the enlargement zone and slightly before the first peak in the thickening zone; and the second peak occurred 40 d later, in the first half of July, when the number of cells in the enlargement zone was decreasing and at the beginning of the second peak in the thickening zone. In contrast, the entry rate into the mature zone followed a different pattern (Fig. 4J–L), reaching a clear maximum at the beginning of July, at the time of the small dip in the number of cells in the thickening zone, and then decreasing but remaining fairly high, before reaching a second peak at the end of the season (October), at the same time as the strong final decrease in the number of cells in the thickening zone.

Discussion

Modelling wood formation dynamics using GFs, GLMs, and GAMs

Numerous studies on the intra-annual dynamics of xylogenesis have been conducted during the last decade, to gain a better understanding of the influence of climate change on wood formation (Gričar *et al.*, 2011). The main method used to investigate xylogenesis involves repeated sampling of the stem through a growing season (Rossi *et al.*, 2006*a*). However, growth is subject to large and sudden variations along and around the stem (Wodzicki and Zajaczkowski, 1970), which makes it difficult to decide whether differences between samples represent the temporal changes in cambial activity (the

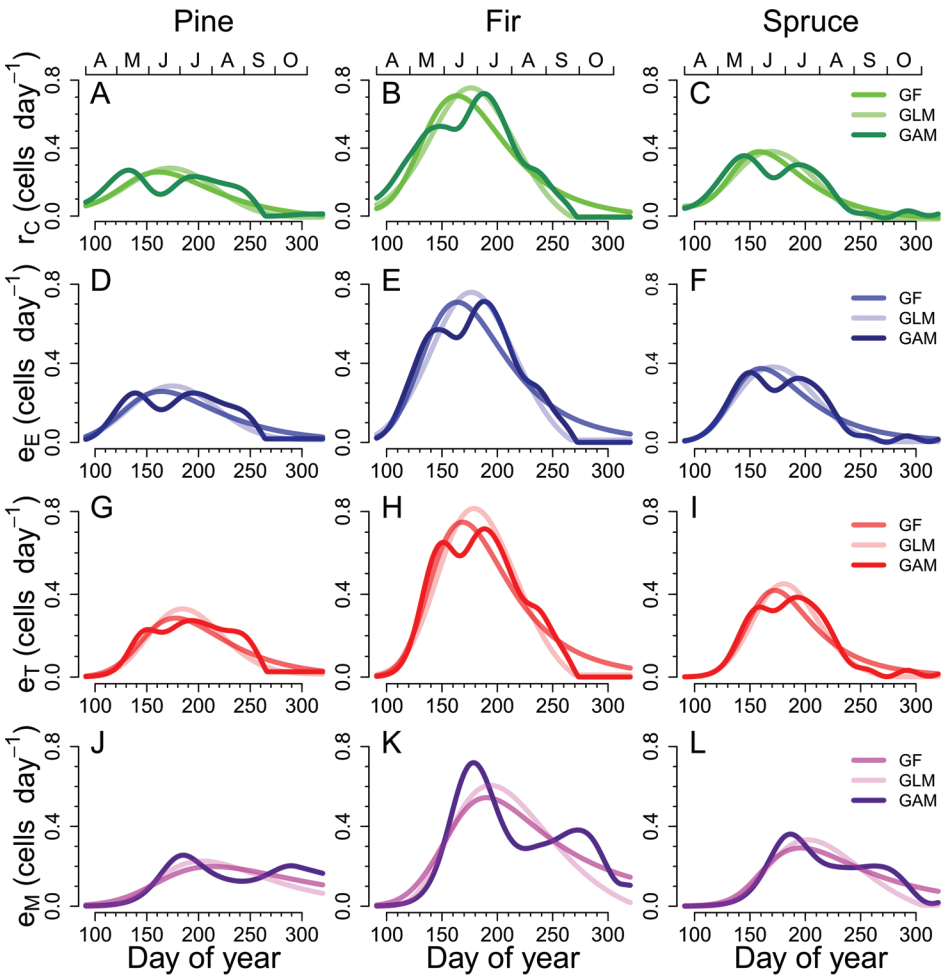

Fig. 4. Rate of cell production (r_C) by the cambial zone, and rates of entrance into the enlargement (e_E), thickening (e_T), and mature (e_M) zones of wood formation as computed from Gompertz functions (GFs), and generalized linear and additive models (GLMs and GAMs). The upper *x*-axis represents the months.

biological signal of interest) or random within-tree particularities ('noise'). To enhance the desired biological signal, Wodzicki (1971) sampled a large number of trees (60 in 1967, 216 in 1968) and obtained excellent results on the kinetics of tracheid development. However, such large sampling is exceptional considering the amount of work and cost involved.

In this study, three statistical approaches (GFs, GLMs, and GAMs) were tested to extract a meaningful biological signal from naturally noisy data sets (groups of five trees from three species over 3 years). It was found that GFs and GLMs accurately represented the general changes in the number of cells in the enlargement and mature zones, but were inaccurate concerning the cambial and thickening zones. Therefore, it is argued that GFs and GLMs could be used to compare general growth patterns between groups of trees—GFs can be used to extract meaningful biological parameters such as the final number of produced cells and the value and occurrence of the maximal rate of cell production (see, for example, Cuny *et al*. 2012)—but it is argued that they are unsuitable for accurately characterizing intra-annual dynamics of the wood formation process (zone entry rates and/or cell timings).

In contrast, GAMs captured in detail the intrinsic complexity of intra-annual wood formation dynamics, making them the only sound approach that is currently available for accurately characterizing this process. Their use solves the basic but thorny issues of sampling date heterogeneity, sampling frequency variability, and missing data. GAMs allow trees and groups of trees (e.g. across species, sites, or years) to be compared or pooled. One possible problem, however, is that GAMs can produce results that are 'biologically incoherent' (e.g. negative zone entry rates), in which case the degree of smoothing must be increased until coherent results are obtained (Wood, 2006).

Because of their high flexibility, GAMs could be adapted to a wide range of conditions where GFs are inappropriate, for example accounting for an additional resting period during the summer drought period in Mediterranean regions (De Luis *et al.*, 2007; Camarero *et al*., 2010), or fitting more complex growth patterns as found in ring-porous deciduous trees (Michelot *et al*., 2012). Moreover, mixed-effect generalized additive models (GAMMs) could accurately account for the non-independence of weekly cell count data (Zuur *et al*., 2009), providing a more robust description of the system.

Intra-annual wood formation dynamics

Intra-annual wood formation dynamics is generally described as three delayed bell-shaped curves for the numbers of cells in the cambial (CZ), enlargement (EZ), and thickening (TZ) zones, followed by an S-shaped curve for the mature (MZ) zone (Rossi *et al*., 2006*a*; Lupi *et al*., 2010; Cuny *et al*., 2012). However, this traditional view, which has been shaped by the use of GFs, did not stand the test of a GAM approach. Rather, it was demonstrated that, for conifers in temperate forests, the general pattern of intra-annual wood formation dynamics is composed of two slightly delayed left-skewed bell-shaped curves representing CZ and EZ, followed by a right-skewed bimodal curve representing TZ, and a double sigmoid curve representing MZ.

The extensive data set of Wodzicki's seminal work (1971) was used to plot the changes in intra-annual cell numbers in the enlargement, thickening, and mature zones for Scots pine trees grown in Poland (Supplementary Fig. S3 at *JXB* online). The patterns thus established were in full agreement with those highlighted in this study, with the curve for enlarging cells being clearly shifted to the left and culminating early (mid-May), while the curve for thickening cells reached a plateau (June) and then increased again to reach a late maximum (September), and the accumulation of mature cells clearly followed a double sigmoid curve. In the 1960s, some authors also observed that the number of cells culminated early in the season for enlargement and late for thickening (Wodzicki and Peda, 1963; Whitmore and Zahner, 1966), which is supported by the present findings.

The calculations showed that the enlargement duration (d_E) decreased from 2–3 weeks to <1 week as the season progressed, while thickening duration (d_T) increased from 3 weeks to nearly 2 months. These values match those of Wodzicki (1971), whereas Horacek *et al*. (1999) found longer d_E but similar d_T. Both Wodzicki (1971) and Horacek *et al*. (1999) also described the same characteristic changes in d_T: steady minimal durations for the first 40% of cells in a ring, a strong increase for the following 40%, and a new plateau around the maximum duration for the last 20%. Conversely, studies based on GFs have found pseudo-linear variations in the differentiation durations between the successive cells of a ring (Deslauriers *et al*., 2003; Rossi *et al*., 2006*a*) and, moreover, have calculated shorter d_T, from 1–3 weeks for the first tracheids to 4–5 weeks for the last tracheids. It should be noted that these studies were performed in colder environments (boreal or subalpine zone), which may explain the discrepancies, but pseudo-linear variations and underestimated values of d_T were also obtained by using GFs.

The developing xylem is a dynamic biological system

The present results show that the developing xylem is a dynamic biological system, justifying the analogy made with hydrological system modelling. It has been demonstrated that the cell flow through the different zones of xylogenesis is controlled not only by the production rate of the cambium (the source), but also by the cell residence times in each 'reservoir' (Supplementary Fig. S4A at *JXB* online), creating the complex pattern described above (Supplementary Fig. S4B), and resulting in strong differences in the computation of the timing of cell development (Supplementary Fig. S4C). The early culmination of cells in the enlargement reservoir can be related to both (i) the high flow from the source (demonstrated by the high division rate in the cambium and the high entry rates into the enlarging reservoir); and (ii) the long cell residence times in this reservoir at the beginning of the season. The enlargement reservoir then progressively emptied, because of the decreasing residence times in this reservoir.

The early peak in cells in the thickening reservoir can also be related to the intense flow from the source at the beginning of the growing season: the many cells born in the cambium

passed quite rapidly through the enlarging reservoir, before entering the thickening reservoir, where they accumulated because of structurally longer residence times. After this first peak, the number of cells in the thickening reservoir remained quasi-stable for a while, indicating that the entry and exit rates were at equilibrium. At the end of summer, despite a rapid decrease in the entry rate, the thickening reservoir reached its maximum cell number, because of an abrupt increase in the cell residence times.

The complex dynamics of the thickening reservoir was demonstrated by the accumulation of mature cells, which had two periods of rapid increase separated by a slower transition period. The first rapid accumulation echoed the intense cell flow generated by the cambium in the first part of the season, which cascaded through the successive reservoirs. In summer, the period of slower accumulation corresponded to the increasing residence times in the thickening reservoir, resulting in a diminution of mature reservoir entry rate. In autumn, the last period of rapid accumulation of mature cells occurred during the final rapid emptying of the thickening reservoir

Relationships between the intra-annual dynamics of xylogenesis and wood anatomy

Conifer tree rings can be divided into two parts: earlywood, containing large-diameter, thin-walled tracheids produced at the beginning of the growing season; and latewood, containing narrow-diameter, thick-walled tracheids produced towards the end of the season. In the present data set, the first latewood cell—identified according to Mork's criterion (Denne, 1988)—was estimated to enter the thickening zone around the beginning of the second peak in thickening cells. Therefore, the bimodal curve describing the number of cells in the thickening zone can be interpreted as two juxtaposed bell-shaped curves, corresponding to two distinct populations of tracheids (earlywood and latewood) exhibiting contrasting behaviours. The higher proportion of latewood in pine (45%) compared with spruce (30%) and fir (27%) explains the greater difference in amplitude between the two peaks in the pine thickening zone.

The observed patterns for cell differentiation durations are consistent with the anatomical profiles of conifer tree rings (Mäkinen *et al*., 2003; Park and Spiecker, 2005; Rathgeber *et al*., 2006). Indeed, the final diameter of a tracheid, which is acquired during the enlargement phase of its differentiation, progressively decreases along a ring, mimicking the changes in d_E; and the final wall thickness, which is acquired during the thickening phase, remains around a steady minimum in the first part of a ring, increases in the middle part, and reaches a steady maximum in the last part, mimicking the changes in d_T. Moreover, the sizes of tracheids are proportional to the time they spent in the different zones of differentiation: in earlywood tracheids, diameters are two to three times larger and d_E two to three times longer than in latewood tracheids; whereas in latewood tracheids, walls are two to three times thicker and d_T two to three times longer than in earlywood tracheids. It is also striking that the changes in d_F along a ring strongly resembled a classic intra-ring wood density profile (Rathgeber *et al*., 2006).

Conclusion

The present study shows that GAMs provide an improved approach for modelling intra-annual wood formation dynamics, which is generally represented by changes in the number of cells in the different development stages of xylogenesis: cambial (CZ), enlargement (EZ), wall thickening (TZ), and mature (MZ) zones. Indeed, only GAMs captured the complexity of intra-annual wood formation dynamics, making them suitable for further characterizing this dynamics by computing both the cell entry rates and the cell residence times in the different zones.

It is argued that the developing xylem must be seen as a dynamic biological system in which the cell production rate interplays with the cell residence time in each zone, resulting in complex intra-annual patterns: two left-skewed bell-shaped curves for CZ and EZ, followed by a right-skewed bimodal curve for TZ and a double sigmoid curve for MZ. These patterns have the advantage of truly reflecting the data and explaining some anatomical features of a tree ring, separating, for example, earlywood and latewood into two distinct cell populations.

It is believed that this approach has great potential and will help to better understand wood formation. It allows a fine assessment of the dynamics of the processes that can be compared directly with observations; for example, a comparison of cell residence times in each of the zones with tracheid anatomical characteristics shows that the change in cell size along a ring is closely related to the change in d_E through the growing season, whereas the change in cell wall thickness is closely related to the change in d_T.

Supplementarydata

Supplementary data are available at *JXB* online.

Figure S1. Classical view of the wood formation dynamics.

Figure S2. Sums of the cell numbers in the wood formation zones as simulated by the models.

Figure S3. Data used by Wodzicki (1971).

Figure S4. Updated view of the wood formation dynamics.

Table S1. Characteristics of the monitored trees.

Table S2. List of the variables.

Table S3. Cell numbers and their variations in the zones of xylogenesis.

Acknowledgements

We thank E. Cornu, E. Farré, C. Freyburger, P. Gelhaye, and A. Mercanti for fieldwork and monitoring; M. Harroué for sample preparation in the laboratory; P. Santenoise for his help with the statistics; and M. Dassot for his help with editing the figures.

References

Abe H, Funada R, Ohtani J, Fukazawa K. 1997. Changes in the arrangement of cellulose microfibrils associated with the cessation

of cell expansion in tracheids. *Trees-Structure and Function* **11,** 328–332.

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, eds. *2nd International Symposium on Information Theory* . Akademiai Kiado, Budapest, 267–281.

Armstrong JS, Collopy F. 1992. Eror measures for generalizing about forecasting methods: empirical comparions. *International Journal of Forecasting* **8,** 69–80.

Bates D, Chambers J. 1992. Nonlinear models. In: Chambers J, Hastie T, eds. *Statistical models in S* . Pacific Grove, CA: Wadsworth and Brooks/Cole, 421–453.

Camarero JJ, Guerrero-Campo J, Gutierrez E. 1998. Tree-ring growth and structure of *Pinus uncinata* and *Pinus sylvestris* in the Central Spanish Pyrenees. *Arctic and Alpine Research* **30,** 1–10.

Camarero JJ, Olano JM, Parras A. 2010. Plastic bimodal xylogenesis in conifers from continental Mediterranean climates. *New Phytologist* **185,** 471–480.

Chaffey N. 2002. Why is there so little research into the cell biology of the secondary vascular system of trees? *New Phytologist* **153,** 213–223.

Cuny HE, Rathgeber CBK, Lebourgeois F, Fortin M, Fournier M, 2012. Life strategies in intra-annual dynamics of wood formation: example of three conifer species in a temperate forest in north-east France. *Tree Physiology* **32,** 612–625.

De Luis M, Gričar J, Čufar K, Raventós J. 2007. Seasonal dynamics of wood formation in *Pinus halepensis* from dry and semi-arid ecosystems in Spain. *IAWA Journal* **28,** 389–404.

Denne MP. 1988. Definition of latewood according to Mork (1928). *IAWA Bulletin* **10,** 59–62.

Deslauriers A, Morin H, Begin Y. 2003. Cellular phenology of annual ring formation of *Abies balsamea* in the Quebec boreal forest (Canada). *Canadian Journal of Forest Research – Revue Canadienne De Recherche Forestière* **33,** 190–200.

Gričar J, Rathgeber CBK, Fonti P. 2011. Monitoring seasonal dynamics of wood formation. *Dendrochronologia* **29,** 123–125.

Guisan A, Edwards TC, Hastie T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* **157,** 89–100.

Hastie T, Tibshirani R. 1986. Generalized additive models. *Statistical Science* **1,** 297–318.

Horacek P, Slezingerova J, Gandelova L. 1999. Effects of environment on the xylogenesis of Norway spruce (*Picea abies* L. Karst.). In: Wimmer R, Vetter RE, eds. *Tree-ring analysis: biological, methodological and environmental aspects* . Wallingford, UK: CABI Publishing, 33–53.

Kutscha NP, Hyland F, Schwarzmann JM. 1975. Certain seasonal changes in balsam fir cambium and its derivatives. *Wood Science and Technology* **9,** 175–188.

Larson PR. 1994. *The vascular cambium: development and structure*. Berlin: Springer-Verlag.

Lupi C, Morin H, Deslauriers A, Rossi S. 2010. Xylem phenology and wood production: resolving the chicken-or-egg dilemma. *Plant, Cell and Environment* **33,** 1721–1730.

Mäkinen H, Nöjd P, Saranpää P. 2003. Seasonal changes in stem radius and production of new tracheids in Norway spruce. *Tree Physiology* **23,** 959–968.

Mayer DG, Butler DG. 1993. Statistical validation. *Ecological Modelling* **68,** 21–32.

McCullagh P, Nelder JA. 1983. *Generalized linear models* . London: Chapman and Hall.

Michelot A, Simard S, Rathgeber C, Dufrêne E, Damesin C. 2012. Comparing the intra-annual wood formation of three European species (*Fagus sylvatica, Quercus petraea* and *Pinus sylvestris*) as related to leaf phenology and non-structural carbohydrate dynamics. *Tree Physiology* **32,** 1033–1045.

Park Y-I, Spiecker H. 2005. Variations in the tree-ring structure of Norway spruce (*Picea abies*) under contrasting climates. *Dendrochronologia* **23,** 93–104.

Plomion C, Leprovost G, Stokes A. 2001. Wood formation in trees. *Plant Physiology* **127,** 1513–1523.

R Development Core Team. 2011. *R: a language and environment for statistical computing* . Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Rathgeber CBK. 2012. Cambial activity and wood formation: data manipulation, visualisation and analysis using R. R package version 1.4-1. http://CRAN.R-project.org/package=CAVIAR.

Rathgeber CBK, Decoux V, Leban J-M. 2006. Linking intra-tree-ring wood density variations and tracheid anatomical characteristics in Douglas fir (*Pseudotsuga menziesii* (Mirb.) Franco). *Annals of Forest Science* **63,** 699–706.

Rathgeber CBK, Rossi S, Bontemps J-D. 2011. Cambial activity related to tree size in a mature silver-fir plantation. *Annals of Botany* **108,** 429–438.

Rossi S, Deslauriers A, Anfodillo T. 2006*a*. Assessment of cambial activity and xylogenesis by microsampling tree species: an example at the alpine timberline. *IAWA Journal* **27,** 383–394.

Rossi S, Deslauriers A, Anfodillo T, Morin H, Saracino A, Motta R, Borghetti M. 2006*b*. Conifers in cold environments synchronize maximum growth rate of tree-ring formation with day length. *New Phytologist* **170,** 301–310.

Rossi S, Deslauriers A, Gričar J, Seo J-W, Rathgeber CBK, Anfodillo T, Morin H, Levanic T, Oven P, Jalkanen R. 2008. Critical temperatures for xylogenesis in conifers of cold climates. *Global Ecology and Biogeography* **17,** 696–707.

Rossi S, Deslauriers A, Morin H. 2003. Application of the Gompertz equation for the study of xylem cell development. *Dendrochronologia* **21,** 33–39.

Rossi S, Morin H, Deslauriers A, Plourde P-Y. 2011. Predicting xylem phenology in black spruce under climate warming. *Global Change Biology* **17,** 614–625.

Skene DS. 1969. The period of time taken by cambial derivatives to grow and differentiate into tracheids in *Pinus radiata* D. Don. *Annals of Botany* **33,** 253–262.

Skene DS. 1972. The kinetics of tracheid development in *Tsuga canadensis* Carr. and its relation to tree vigour. *Annals of Botany* **36,** 179–187.

Vaganov EA, Anchukaitis KJ, Evans MN. 2011. How well understood are the processes that create dendroclimatic records?

A mechanistic model of the climatic control on conifer tree-ring growth dynamics. In: Hughes MK, Swetnam TW, Diaz HF, eds. *Dendroclimatology* . London: Springer-Verlag, 37–75.

Whitmore FW, Zahner R. 1966. Development of the xylem ring in stems of young red pine trees. *Forest Science* **12,** 198–210.

Willmott CJ, Matsuura K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* **30,** 79–82.

Wilson BF, Wodzicki TJ, Zahner R. 1966. Differentiation of cambial derivatives: proposed terminology. *Forest Science* **12,** 438–440.

Wodzicki TJ. 1971. Mechanism of xylem differentiation in *Pinus sylvestris* L. *Journal of Experimental Botany* **22,** 670–687.

Wodzicki TJ, Peda T. 1963. Investigation on the annual ring of wood formation in European silver fir (*Abies pectinata* D.C.). *Acta Societatis Botanicorum Poloniae* **32,** 609–618.

Wodzicki TJ, Zajaczkowski S. 1970. Methodical problems in studies on seasonal production of cambial xylem derivatives. *Acta Societatis Botanicorum Poloniae* **39,** 509–520.

Wood SN. 2006. *Generalized additive models: an introduction with R*. Boca Raton, FL: Chapman and Hall/CRC.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. 2009. *Mixed effects models and extensions in ecology with R* . New York: Springer-Verlag.